Somatic POLE exonuclease domain mutations are early events in sporadic endometrial and colorectal carcinogenesis, determining driver mutational landscape, clonal neoantigen burden and immune response

Temko, Daniel; Van Gool, Inge C; Rayner, Emily; Glaire, Mark; Makino, Seiko; Brown, Matthew; Chegwidden, Laura; Palles, Claire; Depreeuw, Jeroen; Beggs, Andrew; Stathopoulou, Chaido; Mason, John; Baker, Ann-Marie; Williams, Marc; Cerundolo, Vincenzo; Rei, Margarida; Taylor, Jenny C; Schuh, Anna; Ahmed, Ahmed; Amant, Frédéric

DOI:
10.1002/path.5081

Somatic *POLE* exonuclease domain mutations are early events in sporadic endometrial and colorectal carcinogenesis, determining driver mutational landscape, clonal neoantigen burden and immune response

Daniel Temko[1,2,3], Inge C Van Gool[4], Emily Rayner[5], Mark Glaire[5], Seiko Makino[5], Matthew Brown[5], Laura Chegwidden[5], Claire Palles[5], Jeroen Depreeuw[6,7,8], Andrew Beggs[9], Chaido Stathopoulou[5], John Mason[5], Ann-Marie Baker[1], Marc Williams[1,10], Vincenzo Cerundolo[11], Margarida Rei[11], Jenny C Taylor[5], Anna Schuh[12], Ahmed Ahmed[13], Frédéric Amant[14], Diether Lambrechts[7,8], Vincent THBM Smit[4], Tjalling Bosse[4], Trevor A Graham[1], David N Church[5]*, Ian Tomlinson[9]*

AFFILIATIONS

1 Evolution and Cancer Laboratory, Barts Cancer Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK.

2 Centre for Maths and Physics in the Life Sciences and Experimental Biology (CoMPLEX), University College London, London, UK

3 Department of Computer Science, University College London, London, UK

4 Department of Pathology, Leiden University Medical Center, Leiden, the Netherlands.

5 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

6 KU Leuven (University of Leuven), University Hospitals Leuven, Department of Obstetrics and Gynecology, Division of Gynecologic Oncology, Leuven, Belgium

7 KU Leuven, Department of Human Genetics, Laboratory for Translational Genetics, Leuven, Belgium

8 VIB Center for Cancer Biology, Laboratory for Translational Genetics, Leuven, Belgium

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/path.5081

9 Institute of Cancer and Genomic Sciences, University of Birmingham, Edgbaston, Birmingham, UK

10 Department of Cell and Developmental Biology, University College London, London, UK

11 MRC Human Immunology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK

12 Department of Oncology, Old Road Campus Research Building, University of Oxford, Oxford, UK

13 Ovarian Cancer Cell Laboratory, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK; Nuffield Department of Obstetrics & Gynaecology, University of Oxford, Oxford, UK

14 Center for Gynaecologic Oncology Amsterdam, Netherlands Cancer Institute, Amsterdam, The Netherlands.

** Correspondence to David Church, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. E-mail: dchurch@well.ox.ac.uk; Or Prof. Ian Tomlinson, Institute of Cancer and Genomic Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. E-mail: i.tomlinson@bham.ac.uk*

Conflicts of interest statement: All – none to declare.

Running title: Somatic *POLE* mutations are early events in sporadic cancers

Word count: Abstract: 223 words Manuscript (including Methods): 3966 words

ABSTRACT

Genomic instability, a hallmark of cancer, is generally thought to occur in the mid to late stages of tumorigenesis, following the acquisition of permissive molecular aberrations such as *TP53* mutation or whole genome doubling. Tumours with somatic *POLE* exonuclease domain mutations are notable for their extreme genomic instability (their mutation burden is among the highest in human cancer), distinct mutational signature, lymphocytic infiltrate and excellent prognosis. To what extent these characteristics are determined by the timing of *POLE* mutations in oncogenesis is unknown. Here, we have shown that pathogenic *POLE* mutations are detectable in non-malignant precursors of endometrial and colorectal cancer. Using genome and exome sequencing, we found that multiple driver mutations in *POLE*-mutant cancers display the characteristic *POLE* mutational signature, including those in genes conventionally regarded as initiators of tumorigenesis. In *POLE*-mutant cancers, the proportion of monoclonal predicted neoantigens was similar to other cancers, but the absolute number was much greater. We also found that the prominent $CD8^+$T cell infiltrate present in *POLE*-mutant cancers was evident in their precursor lesions. Collectively, these data indicate that somatic *POLE* mutations are an early, quite possibly initiating, event in the endometrial and colorectal cancers in which they occur. The resulting early onset of genomic instability may account for the striking immune response and excellent prognosis of these tumours, as well as their early presentation.

Keywords

POLE, polymerase proofreading, mutation, endometrial cancer, colorectal cancer, precursor lesion

INTRODUCTION

Next generation sequencing (NGS) technologies have, hugely, advanced our understanding of the mechanisms of tumorigenesis. The ability to analyse the entire genome or exome at depth in large numbers of tumours has substantially increased the list of driver genes – that is those which, when mutated, promote tumour growth. It has also revealed that such driver mutations are not always present in the dominant tumour clone [1,2]. This is clinically relevant, because targeting subclonal drivers is likely to kill only a subpopulation of tumour cells, while successful targeting of clonal variants may lead to tumour eradication. Thus, differentiating early, clonal mutations from late, subclonal ones may not only increase our understanding of the mechanisms of oncogenesis, but also inform the clinical management of patients [2].

Fundamentally, all mutations are caused in part by a failure to recognise or repair defects in DNA sequence or chromosome structure. In many cancers, this is a consequence of specific defects in the cellular processes responsible for maintaining genomic integrity [3]. One recently described example is the genomic instability caused by missense mutations in the exonuclease (proofreading) domains of the major replicative DNA polymerases *POLE* and *POLD1* [4]. Polymerase proofreading recognises and corrects mispaired bases incorporated during DNA replication; its perturbation as a result of these mutations is associated with an exceptional number of SNVs (though not indels), and a distinct mutational signature typified by C:G→A:T transversions where the mutated cytosine is in the context TCT, and C:G→T:A transitions where the mutated cytosine is in the context TCG [4-6]. *POLE* and *POLD1* exonuclease domain mutations may occur in the germline, where they cause polymerase proofreading-associated polyposis (PPAP) – a condition characterised by intestinal polyposis and tumours of the colorectum and uterus, among other organs [7]. Somatic *POLE*

exonuclease domain mutations (hereafter simply referred to as *POLE* mutations) occur in sporadic tumours of the endometrium (7-15% cases) [8,9], colorectum (1-2%) [10,11], and less commonly in other cancers (although for reasons that are unclear, somatic *POLD1* exonuclease domain mutations are very uncommon). *POLE*-mutant colorectal and endometrial cancers have an excellent prognosis [8,11-13], probably owing to a robust anti-tumour immune response against the multitude of immunogenic neoantigens they are predicted to harbour [11,14,15]. Very recent reports also suggest that these tumours may be highly responsive to immune checkpoint inhibition [16].

While it is clear that somatic *POLE* mutation causes a mutator phenotype [17] and acts as a cancer driver [4,5], several questions about its contribution to tumorigenesis remain unanswered. One of the most important of these relates to the timing of these mutations in cancer development. If *POLE* mutations are late events, their consequences may be restricted to a subclone of tumour cells, the targeting of which may fail to alter meaningfully tumour behaviour. In contrast, if *POLE* mutations occur early, they could rapidly cause a large number of clonal alterations that may alter prognosis or response to therapy. This is particularly pertinent in the light of recent data suggesting that long-term benefit from immune checkpoint inhibition is limited to patients whose cancers harbour neoantigens in the dominant tumour clone [18]. In contrast to germline mutations in DNA repair pathways in rare inherited syndromes (such as the mismatch repair gene variants that cause Lynch syndrome), the acquisition of genomic instability in sporadic cancers has largely been believed to be a mid- to late-stage event during carcinogenesis [19]. For example, in sporadic colorectal cancer – a tumour type in which the molecular progression of pre-cancers (adenomas) to invasive carcinomas has been well characterised – mismatch repair deficiency (MMR-D) or chromosomal instability (CIN) occur after initiating (epi)mutations in *APC*,

BRAF or *KRAS*, or other events such as whole genome doubling or loss of chromosome 18q [19-24]. Thus, in addition to its clinical relevance, the demonstration that the *POLE* mutator phenotype operates from the first stages of tumour initiation would also reveal a novel pathway of sporadic tumorigenesis. A recent case report of a pathogenic *POLE* mutation in a endometrial cancer and its precursor [25] suggests that these mutations may occur early in tumour development, but the single case precludes generalization of this result.

In this study, we comprehensively examined the timing of pathogenic somatic *POLE* exonuclease domain mutations in sporadic endometrial and colorectal cancers using tumour whole genome sequencing (WGS), public sequencing data from The Cancer Genome Atlas (TCGA) [8,10], and targeted sequencing of additional cohorts of cancers and pre-cancers.

Materials and methods

Ethical approval

Patient consent for research on tumour tissue was obtained at the recruiting centres under local ethical approval. Molecular analysis of anonymised tissue was performed under Oxford Research Ethics Committee A approval (05/Q1605/66).

Patients and tumour samples

Details of the cohorts and cases analysed in this study are shown in supplementary material, Tables S1 and S2. Fifty one formalin-fixed paraffin-embedded (FFPE) endometrial cancers carrying known pathogenic somatic *POLE* exonuclease domain mutations identified in our previous studies [12,14,26] were reviewed for the presence of a concomitant and spatially discrete area of endometrial intraepithelial neoplasia (EIN) by examination of haematoxylin and eosin (H&E) stained slides by two expert gynaecological pathologists (VS & TB). An additional 389 FFPE colorectal polyps (tubular adenomas, tubulovillous adenomas and serrated adenomas – hereafter referred to as adenomas), for which *POLE* screening had not previously been performed, were identified from 261 participants in the CORGI study, which recruited patients with a family history of colorectal cancer and a personal history of a colorectal polyp or colorectal malignancy in the absence of a known tumour predisposition syndrome. Six fresh frozen tumours with pathogenic somatic *POLE* mutations (five endometrial, one colorectal) were identified from a Leuven endometrial cancer cohort used in our previous study [12], a prospective clinical sequencing programme (HICF2) at the University of Oxford, or the University of Birmingham tissue bank. TCGA colorectal (COADREAD) [10] and endometrial (uterine corpus endometrial carcinoma – UCEC) [8]

cancer data were downloaded from the Genomic Data Commons (GDC) Data Portal (https://portal.gdc.cancer.gov; June 2017). An additional series of 78 FFPE endometrial cancers including 32 cases with pathogenic somatic *POLE* mutations were identified from the LUMC archives (2001-2015) [14]. Further details of the cohorts used in this study are provided in supplementary material, Table S1. Molecular analyses were performed on a single tumour or precursor lesion region in each case.

DNA extraction

After review to confirm adequate tumour cellularity, DNA was extracted from fresh frozen or microdissected FFPE tumours and precursors using standard methods (Roche FFPE-T DNA kit (F. Hoffman La Roche AG, Basel, Switzerland), Machery Nagel Nucleospin DNA FFPE XS (Machery Nagel, Duren, Germany)/ FFPE DNA kit or Qiagen Blood and Tissue kit (Qiagen, Hilden, Germany) and resuspended in buffer or water.

DNA sequencing

Full details of the sample preparation and the sequencing methods utilized in this study are provided in supplementary material, Supplementary materials and methods. In brief, endometrial epithelial neoplasias (EIN) and paired carcinomas were sequenced for mutations in 30 cancer genes using molecular inversion probe capture, and a custom version of the 72 gene Ion AmpliSeq Cancer Hotspot panel v2 (including 80 genes; ThermoFisher, MA, USA) (supplementary material, Tables S3,4). Whole genome sequencing (WGS) of fresh frozen

tumours was performed by Illumina HiSeq (Illumina, San Diego, CA, USA), and aligned to the reference genome by BWA mem or Isaac [27]. FFPE endometrial cancers from the LUMC series were analysed using the Lifetech/ThermoFisher Ion AmpliSeq Comprehensive Cancer Panel comprising 409 cancer genes (http://www.lifetechnologies.com/order/catalog/product/4477685). Mutation calling was performed by LoFreq [28] (EINs), Mutect, Mutect2 [29] or Strelka [30] (WGS, TCGA cases), or Ion Torrent variantCaller (EINs, LUMC FFPE tumours). Copy number profiles were derived using Sequenza [31]. Variant annotation was done using Annovar [32] or Variant Effect Predictor [33].

Definition of driver genes

Driver genes were defined using the IntOGen driver gene repository (https://www.intogen.org/search) and included both PanCancer (Pooled_driver) and tumour type-specific (perProject_driver) variants (supplementary material, Tables S5, S6) [34]. High confidence driver mutations (defined as either truncating mutations in genes likely to be tumour suppressors or recurrent missense mutations in any endometrial or colorectal cancer-specific or pan-cancer gene from the IntOGen set) were determined for a subset of driver genes by manual curation, blinded to tumour molecular characteristics.

Clonality of POLE mutations

Most (36 of 38) endometrial and colorectal cancers with pathogenic *POLE* mutations were disomic at the *POLE* locus (chr12q24) and were informative for clonality analysis. Of these, 20 of 22 endometrial cancers, and 12 or 14 colorectal cancers had available copy number

annotation. As all 32 of these showed near-diploid genomes (>80% of the genome), we assumed diploid genomes for the four remaining cases.

Mutations were filtered to include only autosomal variants in diploid regions of the genome, called with depth of at least 20x. Mutation allele frequency distributions were generated using the R 'histogram' function, and tumour cellularity inferred as twice the mid-point of the allele frequency bin with highest mutation density, excluding bins with a lower bound below allele frequency 0.1. These values were then subjected to manual curation. The hypothesis that the mutation was present in every tumour cell was tested by a one-sided binomial test, based on the numbers of reference and variant reads at the *POLE* mutation site and the inferred tumour cellularity. Specifically, for a mutation with coverage R, in a tumour with tumour cell fraction C, the number of variant reads was modelled as a random variable X, with distribution:

$$X \sim Binom(R, C / 2).$$

In each case we calculated the probability, p, of finding the observed number of variant reads, v, or fewer, $P(X<=v)$. Mutations were considered subclonal for $p<=0.05$.

Mutational signatures

Previously reported mutational signatures were obtained from http://cancer.sanger.ac.uk/cosmic/signatures/ on 1 June 2017. The complement of mutational processes active in the life-history of each tumour sample was inferred by classification of mutations into 96 categories following Alexandrov [6], and the use of non-negative least

squares regression, implemented in the R package 'nnls'. For this analysis, only mutational signatures previously reported as active in that cancer type (endometrial signatures 1, 2, 5, 6, 10, 13, 14 and 26; colorectal signatures 1, 5, 6, and 10) were used for the regression. For cases analysed by whole exome sequencing, mutational signatures were re-scaled to exomic trinucleotide frequencies. A mutational process was deemed to have been active in the life-history of a tumour if the associated mutational signature had a coefficient of at least 2 per cent of the total coefficients in the best-fitting model. Mutations likely to be due to *POLE* exonuclease domain mutation (*POLE*) were identified by considering mutational signatures as multinomial probability distributions caused by specific mutational processes. The probability of each mutation under all mutational processes active in that tumour was calculated, and mutations were assigned to the "POLE" mutational process in cases where the probability under that process was at least twice the probability under any other process.

POLE consensus mutational signature scores in driver genes

Tumour mutations were obtained from calling based on tumour/normal .bam files (*POLE* mutant cases) or TCGA MAF files (MMR-P, MMR-D cases), and classified into 96 categories following Alexandrov [6]. For each tumour, the distribution of mutations across the 96 types was calculated, and re-scaled to equal trinucleotide frequencies based on sequencing type, thus obtaining an individual tumour mutational signature. Tumours were then categorised into three groups according to *POLE* mutation and mismatch repair status (i.e. *POLE*-mutant, MMR-P and MMR-D), and a consensus mutational signature was calculated for each group as the average of the individual-tumour signatures among samples in the group, weighted by the number of mutations in each sample. The probability of all non-

silent mutations ('nonsynonymous SNV', or 'stopgain') in driver genes (as defined above) under each of the three consensus mutational signatures was then calculated, and the ratio of the probability of each mutation under the *POLE* consensus mutational signature compared to that under each of the other two consensus mutational signatures was obtained. For each individual gene, a '*POLE* score' was then calculated as the base two logarithm of the minimum value of these ratios across all the non-silent mutations within that gene.

Immunohistochemistry

Immunohistochemistry (IHC) for CD8 was performed as reported previously [14]. The number of $CD8^+$ cells was quantified for the epithelial and stromal regions of the EIN. For the final CD8 count per case, the mean of these regions in ten high-power fields (HPF; 625 μm x 425 μm) was calculated. A similar method was used to quantify CD8 density in colorectal adenomas, although the small lesion size meant that estimates were obtained from the mean of two or three HPFs.

Clonal neoantigen prediction

We estimated the number of clonal neoantigens using a modification of our previously-reported algorithm [11], modified to predict peptide binding against patient-specific HLA molecules (determined from WGS or WES data using OptiType [35]). Neoantigens were defined as mutations predicted to specify peptides that bound patient HLA molecules with affinity <500 nM. Copy number information was obtained from the GDC data portal, as

described above. Clonality was determined as described above. Neoantigens were considered clonal if the binomial test *P*-value was over 0.05.

Statistical analysis

Analyses were performed using R (CRAN network) or Prism (GraphPad Software, La Jolla, CA, USA). Statistical comparison between groups was made using the non-parametric Mann-Whitney U test. All *P* values were two sided, unless otherwise specified. Statistical significance was accepted at $P<0.05$.

RESULTS

Somatic POLE *exonuclease domain mutations are detectable in sporadic endometrial and colorectal pre-cancers*

As somatic *POLE* exonuclease domain mutations have been best characterised in endometrial and colorectal cancers, we first examined whether these mutations were present in precursors of these malignancies. Expert histopathological review of 51 *POLE*-mutant endometrial cancers revealed four with a concomitant and spatially discrete area of endometrial intraepithelial neoplasia (EIN), the precursor of endometrioid carcinoma (supplementary material, Table S2). Microdissection and targeted sequencing of these lesions by a 30-gene molecular inversion probe capture NGS panel (supplementary material, Table S3), a custom 80 gene Ion Ampliseq Cancer Hotspot panel (supplementary material, Table S4) and Sanger sequencing revealed that in all cases, the *POLE* mutation present in the carcinoma was also detectable in the paired precursor (Figure 1A,B, supplementary material, Table S7). While some other driver mutations were also shared between the precursors and paired cancers (median 4 shared mutations per pair, relative to a median of 7 mutations per EIN and median of 10 mutations per carcinoma), the progression from EIN to malignancy was associated with both the loss (median 3 mutations lost in carcinomas compared to paired EINs) and, more frequently, gain (median 6 mutations gained in carcinomas compared to paired EINs) of driver mutations (Figure 1A,B, supplementary material, Table S7). Notably, many of the driver mutations gained were replacements of a glutamic acid or arginine codon with a nonsense codon (E→* or R→ *), consistent with the characteristic mutational bias associated with *POLE* mutation (C:G→A:T transversions where the mutated cytosine is in the context

T<u>C</u>T, and C:G→T:A transitions where the mutated cytosine is in the context T<u>C</u>G) [4-6] (Figure 1B, supplementary material, Table S7).

We were unable to perform a corresponding analysis of colorectal tumours, because residual precursor is uncommon in colorectal carcinomas. However, screening of 389 colorectal adenomas from 261 patients revealed three (0.8% adenomas, 1.1% patients) with somatic *POLE* mutations (Figure 1C), a frequency concordant with that found in colorectal cancers [11]. Unfortunately, the limited amount of DNA available from these lesions precluded analysis of other driver mutations.

Mutational landscape and driver gene alterations suggest that somatic POLE *mutation is an early event in sporadic endometrial and colorectal cancers*

To further investigate the timing and consequences of *POLE* mutations in tumour development, we performed WGS on six cancers (five endometrial, one colorectal), all of which harboured the most common pathogenic *POLE* exonuclease domain variant – a proline to arginine substitution at codon 286 (*POLE*^{P286R}) (Figure 2A). Each displayed a substantially elevated mutation burden (122–731 mutations/Mb), and characteristic preponderance of C:G→A:T substitutions in the context T<u>C</u>T (Figure 2A,B, supplementary material, Table S8, Figure S1) [6]. In keeping with their early occurrence, both the *POLE* mutations themselves, and other mutations consistent with the known *POLE* mutational signature (see Materials and methods, Mutational signatures) appeared clonal in all six cases (Figure 2C). This was also the case in 17 of 17 endometrial cancers and 12 of 13 colorectal cancers with pathogenic *POLE* exonuclease domain mutations from the TCGA series (supplementary material,

Figures S2, S3). This analysis showed that *POLE* mutations were unlikely to occur as late events after the most recent common ancestor in cancer evolution.

We next examined the timing of *POLE* mutations in carcinogenesis in more detail by analysis of driver genes, including some known usually to be mutated early in the pathogenesis of endometrial or colorectal cancer. To assess the likelihood that mutations in these genes were secondary to an earlier *POLE* mutation, we developed a metric to score them according to the probability that they were caused by the mutational process dominant in *POLE*-mutant cancers (presumably caused by the *POLE* mutation itself), rather than the mutational processes operative in other tumours (see Materials and methods, *POLE* consensus mutational signature score for details). For this analysis, we combined our cohort of *POLE*-mutant tumours with *POLE*-mutant cases from TCGA, using MMR-P and MMR-D TCGA cases as comparators. Strikingly, in *POLE*-mutant tumours, almost all known cancer driver genes displayed evidence of the *POLE* consensus mutational signature, with the notable exception of *POLE* itself (Figures 3,4, supplementary material, Tables S8–S10, Figures S4, S5), consistent with the postulate that the *POLE* signature is a direct effect of the polymerase proofreading mutation. In contrast, MMR-P and MMR-D tumours rarely showed evidence of the *POLE* consensus mutational signature (Figures 3,4, supplementary material, Tables S8–S10). In total, among 206 endometrial and/or colorectal cancer driver genes examined in the cases from the combined endometrial and colorectal cancer cohorts, 50% (1,065/2,118) of those in *POLE* mutant samples had a *POLE* signature score >0, compared to 14% (628/4,427) in MMR-D and MMR-P cancers ($P<1x10^{-26}$).

To minimise the possibility of confounding by non-pathogenic mutations in the complete set of driver genes, we repeated these analyses considering only manually curated, high-

confidence pathogenic mutations, and obtained similar results ($P<1x10^{-26}$, supplementary material, Figures S6, S7). As mutation of the tumour suppressors *PTEN* and *APC* are well recognised as early, if not initiating, events in the pathogenesis of endometrial and colorectal cancers respectively, we specifically examined whether somatic variants in these genes varied according to tumour *POLE* mutation status. Among high-confidence pathogenic *PTEN* mutations in endometrial cancers, the proportion with *POLE* consensus mutational signature scores >0 was substantially and significantly greater among *POLE*-mutant cases than among MMR-P and MMR-D tumours (10 of 14 [71.4%] versus 14 of 82 [17.1%] mutations respectively; $P=7.8x10^{-3}$, Fisher's Exact Test). Analysis of high-confidence pathogenic *APC* mutations in colorectal cancers revealed similar results (corresponding proportions 9 of 14 [64.3%] versus 10 of 69 [14.5%] mutations; $P=0.012$, Fisher's Exact Test).

Further analysis of these cohorts and of targeted sequencing data from an additional series of endometrial cancers from the Leiden University Medical Centre (LUMC), including 32 *POLE*-mutant tumours, confirmed the over-representation of E→*, R→* and arginine to glutamine substitutions (R→Q) among *POLE*-mutant cases, concordant with the results from the paired endometrial lesions and consistent with the known trinucleotide bias of the *POLE* mutational signature (supplementary material, Figure S8, S9, S10, Tables S7-S11). Interestingly, this was evident not only in well characterised drivers such as *PTEN* in endometrial cancer and *APC* in colorectal cancer as noted above, but also in recurrent, clonal driver mutations found rarely in that tumour type. For example, in the combined TCGA/LUMC endometrial cancer cohorts, truncating mutations in the tumour suppressors *APC, NF1* and *RB1* were very rare in *POLE*-wild-type tumours (1.1%, 1.5% and 1.5% respectively), but common among *POLE*-mutant cases (38.8%, 34.7% and 34.7% respectively; $P<0.001$ for each comparison, Fisher's exact test), where they almost invariably

occurred at glutamic acid or arginine codons (supplementary material, Figure S8, S9, S10, Tables S9, S11).

Collectively, these data suggested that somatic *POLE* mutation occurs early in endometrial and colorectal cancers, and that its attendant mutator phenotype defines a distinct pathway of carcinogenesis from the initial stages of this process.

Somatic POLE *mutations are associated with a prominent T cell infiltrate in both precancerous and cancerous lesions*

Somatic *POLE* exonuclease domain mutations in endometrial and colorectal cancers are associated with enhanced tumour immunogenicity and favourable prognosis [11,14,15]. We speculated that the early acquisition of somatic *POLE* mutations would cause a rapid acquisition of mutations, some of which would produce neoantigens capable of eliciting an anti-tumour immune response. Consistent with this prediction, all *POLE*-mutant EINs displayed a prominent $CD8^+$ infiltrate (Figure 5A), which was significantly greater than that in *POLE*-wild-type EINs (median 59.4 versus 14.8 $CD8^+$ cells per high power field [HPF]; P=0.029 Mann Whitney U test), and exceeded that observed in the *POLE*-wild-type endometrial carcinomas, although this difference was not statistically significant (median 59.4 versus 24.7 $CD8^+$ cells per HPF, P=0.11) (Figure 5B). The increased $CD8^+$ cell density in *POLE*-mutant EINs could not obviously be explained by other factors such as patient age, or the stage or grade of the paired carcinoma (supplementary material, Table S2). In contrast, the differences in $CD8^+$ density between EINs and paired carcinomas among both *POLE*-wild-type and *POLE*-mutant cases were less marked (median 14.8 versus 24.7; P=0.34, and

59.4 versus 116.9; *P*=0.11 respectively). The single *POLE*-mutant colorectal adenoma for which IHC was possible also demonstrated a dense $CD8^+$ infiltrate (154.9 versus median 34.0 $CD8^+$ cells per HPF) (Figure 5A,B).

Somatic POLE *mutations in colorectal cancer are associated with enhanced predicted clonal neoantigen burden*

Recent data have shown that the presence of predicted neoantigens within the major tumour clone correlates with benefit from immune checkpoint inhibitor therapy [18]. As the limited amount of FFPE-derived DNA from precursor lesions was inadequate for clonality analysis and neoantigen prediction, we examined predicted neoantigen clonality in a subset of TCGA colorectal cancers including MMR-P, MMR-D and *POLE*-mutant subtypes, broadly matched for patient age and tumour stage. We used an approach similar to our previous reports [11,14], modified to incorporate patient-specific HLA haplotypes obtained using OptiType [35] and estimates of tumour clonality derived from analysis of variant allele frequencies (See Materials and methods, Clonal neoantigen prediction). Analysis of our combined cohort by this pipeline confirmed that *POLE*-mutant colorectal cancers harboured a substantially greater number and density of predicted clonal neoantigens (0.12 per Mb) than tumours lacking *POLE* mutations, including both MMR-P (0.0029 per Mb; *P*=0.0002, Mann Whitney U test) and hypermutated MMR-D cases (0.044 per Mb; P=0.03) (Figure 6, supplementary material, Figure S11).

DISCUSSION

In this study, we have presented multiple lines of evidence to show that pathogenic, somatic *POLE* exonuclease domain mutations are usually early and as far as we can detect initiating events in endometrial and colorectal tumorigenesis. We show that the acquisition of *POLE* mutation causes a distinct pattern of mutations in cancer driver genes, substantially increased mutation burden and an enhanced immune response, detectable even in precancerous lesions. Furthermore, we show that early somatic *POLE* exonuclease domain mutations are likely to cause an enrichment of clonal neoantigens that may explain their good prognosis and excellent response to immune checkpoint inhibitors.

APC mutation has traditionally been regarded as the initiating event in sporadic colorectal cancers that develop along the canonical pathway [19], while mutation of *PTEN* is thought to play a similar role in sporadic endometrioid endometrial cancers [36]. Our evidence suggests that in sporadic colorectal and endometrial cancers with pathogenic somatic *POLE* mutations, the *POLE* mutation is antecedent to either of these events. The consequent mutator phenotype it causes influences the type of mutations in these genes and that of the other earliest driver mutations in these cancers, as well as determining their overall mutational landscape [6]. Whether any of these *POLE*-induced driver mutations represent targetable alterations will be an important topic for future research. Similarly, while the increased burden of predicted clonal neoantigens in *POLE*-mutant tumours may explain their enhanced immunogenicity, further work is required to understand the molecular factors that determine this and its therapeutic implications. A further intriguing possibility is that the mutator phenotype and mutational bias drives cancers into an evolutionary cul-de-sac of sub-optimal fitness. The presence of *APC* mutations as an alternative to *CTNNB1* mutations in some *POLE*-mutant endometrial cancers is an exemplar, and there are likely to be others, such as *NF1* and *RB1*

mutations in endometrial cancer and atypical (Q61P, K117N and A146T) *KRAS* mutations in colorectal cancer. Examination of this hypothesis by comparing the oncogenic effects of these uncommon mutations with those caused by more typical variants in model systems would be of considerable interest.

Our data add to the expanding body of evidence suggesting that the effects of genomic instability in cancer depend upon both its severity and timing. For example, upregulation of APOBEC cytosine deaminase enzymes is common in many types of cancers, resulting in an increased mutation rate and characteristic mutation spectrum [6]. However, APOBEC overexpression often occurs as a late event in advanced tumours and causes a more modest mutator phenotype than *POLE* mutations [2,6]. Speculatively, these features may explain why the impact of APOBEC on prognosis appears more variable than that of *POLE* mutation [37,38]. The early acquisition of somatic *POLE* mutations in sporadic cancers may also help to explain their association with young age at diagnosis, given the prediction that the early gain of a mutator phenotype will accelerate the process of malignant transformation [39].

Our study has limitations. The number of precursor lesions informative for detailed analysis was limited, in keeping with the relative rarity of *POLE* mutations in endometrial cancer, and the frequency with which precancerous and cancerous lesions occur in the same tumour section. Moreover, although the spatial separation of the precancerous and cancerous compartments, and the discordance in molecular alterations between the two components in each case suggests otherwise, we cannot exclude the possibility that the apparent precursor lesion is in fact adenocarcinoma colonizing endometrial glands. It will therefore be important to validate our results in additional cohorts, although we note that a very recent study has documented a pathogenic *POLE* mutation in an endometrial cancer precursor [25].

Furthermore, all our results are based on the analysis of a single sample of each cancer, meaning that the effects of intratumour heterogeneity on the pattern of driver mutations and clonal neoantigens in *POLE*-mutant tumours requires further definition. However, the absence of multi-region sequencing is unlikely to have confounded the principal conclusions of our study regarding the timing of these pathogenic mutations in cancers.

In summary, we show that pathogenic, somatic *POLE* exonuclease domain mutations are early, quite possibly initiating, events in sporadic cancers, and strongly shape subsequent tumour evolution. Our observation provides further insights into the distinct biology of these tumours, and may help explain their increased immunogenicity and excellent prognosis.

Acknowledgments

We thank Michelle Osse, Natalja ter Haar, Nienke Solleveld-Westerink and Dina Ruano for help with performing and interpretation of the Ion AmpliSeq Cancer Hotspot Panel. We would also like to thank the patients who consented for the molecular analysis their tumours reported in this study. The results published here include data generated by The Cancer Genome Atlas project funded by the NCI and NIH (cancergenome.nih.gov); we would like to thank the many patients and researchers who contributed to this study

Funding

Funding for this study was provided by: Cancer Research UK (C6199/A10417 and C399/A2291), the European Union Seventh Framework Programme (FP7/207- 2013) grant 258236 collaborative project SYSCOL, European Research Council project EVOCAN, the Dutch Cancer Society, Research Fund Flanders (F.W.O.) grant no. G.0827.13, the Medical Research Council, the Wellcome Trust and Department of Health as part of a Health Innovation Challenge Fund grant (R6-388), the Oxford NIHR Comprehensive Biomedical Research Centre, Ovarian Cancer Action and core funding to the Wellcome Trust Centre for Human Genetics from the Wellcome Trust (090532/Z/09/Z). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health or the Wellcome Trust.

DT is funded by an EPSRC doctoral training grant via CoMPLEX. MG is funded by a Studentship from the Wellcome Trust. TB is funded by the Dutch Cancer Society Young

Investigator Grant 10418. FA is a senior researcher for the F.W.O. TG is a Cancer Research UK Career Development Fellow (A19771) and a Wellcome Trust Investigator (202778/Z/16/Z). DNC is funded by a Health Foundation/Academy of Medical Sciences Clinician Scientist Fellowship.

The cost of open access publication were provided by core funding to the Wellcome Trust Centre for Human Genetics from the Wellcome Trust (090532/Z/09/Z).

Author contributions statement

Study design: DT, TG, DNC, IT

Data collection: DT, IVG, ER, SM, MB, LC, CP, JD, AB, CS, JM, VC, MR, AA, FA, DL, VS, TB

Data analysis: DT, IVG, ER, MG, LC, CP, AMB, MW, MR, JT, AS, VS, TB, TG, DNC, IT

Data interpretation: DT, TG, DNC, IT

Manuscript writing: DNC, IT

REFERENCES

1. Kandoth C, McLellan MD, Vandin F*, et al.* Mutational landscape and significance across 12 major cancer types. *Nature* 2013; **502:** 333–339.
2. McGranahan N, Favero F, de Bruin EC*, et al.* Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med* 2015; **7:** 283ra254.
3. Tubbs A, Nussenzweig A. Endogenous DNA damage as a source of genomic instability in cancer. *Cell* 2017; **168:** 644–656.
4. Rayner E, van Gool IC, Palles C*, et al.* A panoply of errors: polymerase proofreading domain mutations in cancer. *Nat Rev Cancer* 2016; **16:** 71–81.
5. Shinbrot E, Henninger EE, Weinhold N*, et al.* Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. *Genome Res* 2014; **24:** 1740–1750.
6. Alexandrov LB, Nik-Zainal S, Wedge DC*, et al.* Signatures of mutational processes in human cancer. *Nature* 2013; **500:** 415–421.
7. Palles C, Cazier JB, Howarth KM*, et al.* Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* 2013; **45:** 136–144.
8. The Cancer Genome Atlas Network. Integrated genomic characterization of endometrial carcinoma. *Nature* 2013; **497:** 67–73.
9. Church DN, Briggs SE, Palles C*, et al.* DNA polymerase epsilon and delta exonuclease domain mutations in endometrial cancer. *Hum Mol Genet* 2013; **22:** 2820–2828.

10. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012; **487:** 330–337.

11. Domingo E, Freeman-Mills L, Rayner E*, et al.* Somatic POLE proofreading domain mutation, immune response, and prognosis in colorectal cancer: a retrospective, pooled biomarker study. *Lancet Gastroenterol Hepatol* 2016; **1:** 207–216.

12. Church DN, Stelloo E, Nout RA*, et al.* Prognostic significance of POLE proofreading mutations in endometrial cancer. *J Natl Cancer Inst* 2015; **107:** 402.

13. Meng B, Hoang LN, McIntyre JB*, et al.* POLE exonuclease domain mutation predicts long progression-free survival in grade 3 endometrioid carcinoma of the endometrium. *Gynecol Oncol* 2014; **134:** 15–19.

14. van Gool IC, Eggink FA, Freeman-Mills L*, et al.* pole proofreading mutations elicit an antitumor immune response in endometrial cancer. *Clin Cancer Res* 2015; **21:** 3347–3355.

15. Howitt BE, Shukla SA, Sholl LM, *et al.* Association of Polymerase e-mutated and microsatellite-instable endometrial cancers with neoantigen load, number of tumor-infiltrating lymphocytes, and expression of PD-1 and PD-L1. *JAMA Oncol* 2015; **1:** 1319–1323.

16. Santin AD, Bellone S, Buza N*, et al.* Regression of chemotherapy-resistant Polymerase epsilon (POLE) ultra-mutated and MSH6 hyper-mutated endometrial tumors with nivolumab. *Clin Cancer Res* 2016; **22:** 5682–5687.

17. Kane DP, Shcherbakova PV. A common cancer-associated DNA polymerase ε mutation causes an exceptionally strong mutator phenotype, indicating fidelity defects distinct from loss of proofreading. *Cancer Res* 2014; **74:** 1895–1901.

18. McGranahan N, Furness AJ, Rosenthal R*, et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* 2016; **351:** 1463–1469.

19. Fearon ER. molecular genetics of colorectal cancer. *Annu Rev Pathol* 2011; **6:** 479–507.

20. Kambara T, Simms LA, Whitehall VL*, et al.* BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum. *Gut* 2004; **53:** 1137–1144.

21. Minoo P, Moyer MP, Jass JR. Role of BRAF-V600E in the serrated pathway of colorectal tumourigenesis. *J Pathol* 2007; **212:** 124–133.

22. Fujiwara T, Bandi M, Nitta M*, et al.* Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* 2005; **437:** 1043–1047.

23. Thompson SL, Compton DA. Proliferation of aneuploid human cells is limited by a p53-dependent mechanism. *J Cell Biol* 2010; **188:** 369–381.

24. Burrell Ra, McClelland SE, Endesfelder D*, et al.* Replication stress links structural and numerical cancer chromosomal instability. *Nature* 2013; **494:** 492–496.

25. Miyamoto T, Ando H, Asaka R*, et al.* Mutation analysis by whole exome sequencing of endometrial hyperplasia and carcinoma in one patient: Abnormalities of polymerase epsilon and the phosphatidylinositol-3 kinase pathway. *J Obstet Gynaecol Res* 2018; **44:** 179–183.

26. Stelloo E, Bosse T, Nout RA*, et al.* Refining prognosis and identifying targetable pathways for high-risk endometrial cancer; a TransPORTEC initiative. *Mod Pathol* 2015; **28:** 836–844.

27. Raczy C, Petrovski R, Saunders CT, *et al.* Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics* 2013; **29:** 2041–2043.

28. Wilm A, Aw PP, Bertrand D, *et al.* LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res* 2012; **40:** 11189–11201.

29. Cibulskis K, Lawrence MS, Carter SL, *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013; **31:** 213–219.

30. Saunders CT, Wong WS, Swamy S, *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 2012; **28:** 1811–1817.

31. Favero F, Joshi T, Marquard AM, *et al.* Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol* 2015; **26:** 64–70.

32. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38:** e164.

33. McLaren W, Gil L, Hunt SE, *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* 2016; **17:** 122.

34. Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nat Methods* 2013; **10:** 1081–1082.

35. Szolek A, Schubert B, Mohr C, *et al.* OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* 2014; **30:** 3310–3316.

36. Mutter GL, Lin MC, Fitzgerald JT, *et al.* Altered PTEN expression as a diagnostic marker for the earliest endometrial precancers. *J Natl Cancer Inst* 2000; **92:** 924–930.

37. Leonard B, Starrett GJ, Maurer MJ, *et al.* APOBEC3G expression correlates with T-cell infiltration and improved clinical outcomes in high-grade serous ovarian carcinoma. *Clin Cancer Res* 2016; **22:** 4746–4755.

38. Sieuwerts AM, Willis S, Burns MB, *et al.* Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. *Horm Cancer* 2014; **5:** 405–413.

39. Beckman RA, Loeb LA. Efficiency of carcinogenesis with and without a mutator mutation. *Proc Natl Acad Sci U S A* 2006; **103:** 14140–14145.

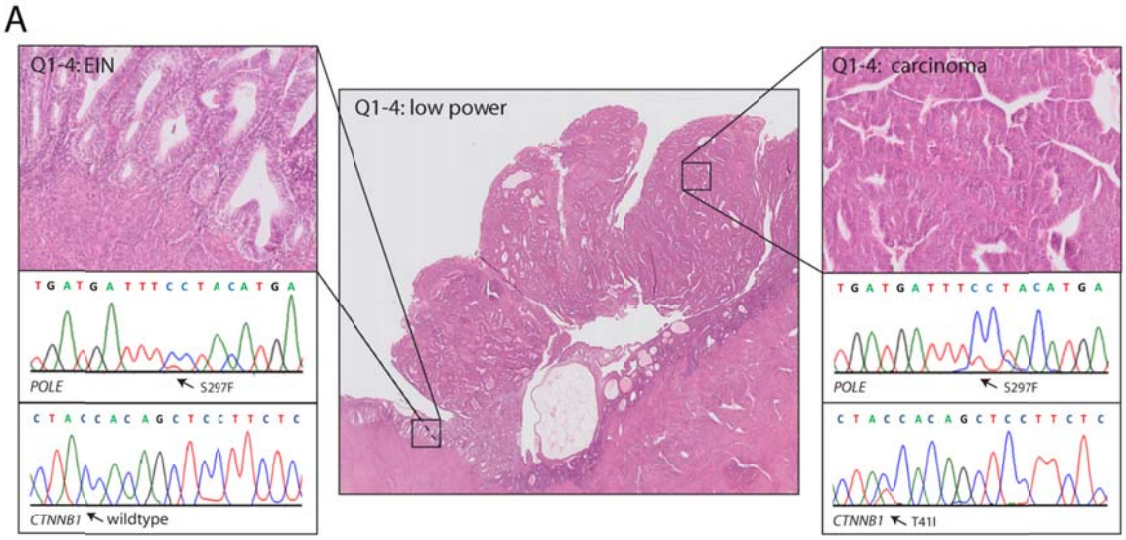
*40. Findlay JM, Castro-Giner F, Makino S, *et al.* Differential clonal evolution in oesophageal cancers in response to neo-adjuvant chemotherapy. *Nat Commun* 2016; **7:** 11111.

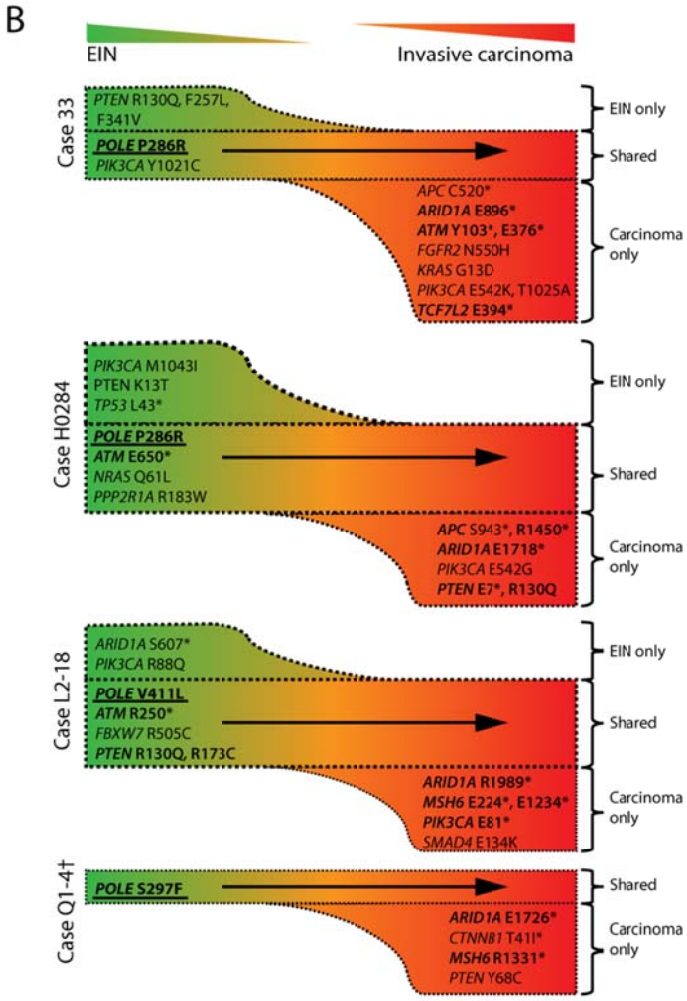
*Cited only in supplementary material.

Figure legends

Figure 1. Pathogenic, somatic *POLE* exonuclease domain mutations in precursors of endometrial and colorectal cancers

Expert histopathological review of 51 endometrial cancers with pathogenic *POLE* exonuclease domain mutations revealed four with concomitant and spatially discrete area of endometrial intraepithelial neoplasia (EIN). (A) H&E stained section from one case with results of Sanger sequencing of the malignant and precursor components. (B) Targeted sequencing of paired endometrial lesions by two orthogonal next generation sequencing panels revealed that *POLE* mutations (bold, underlined) were present in both EIN and carcinomas in all cases (validated by Sanger sequencing in all cases). In each case, progression of EIN to endometrial carcinoma was associated with the gain of driver mutations, several of which were glutamic acid or arginine to stop codon mutations (E→* or R→ *) consistent with the *POLE* exonuclease domain-mutant mutational signature (semibold). †The amount of DNA available from the EIN in case Q1-4 was insufficient for molecular inversion probe sequencing. Details of identified driver mutations are provided in supplementary material, Table S7. (C) H&E stained section from colorectal adenoma with the results of Sanger sequencing and allelic discrimination PCR for the wild-type G allele and mutant T allele.





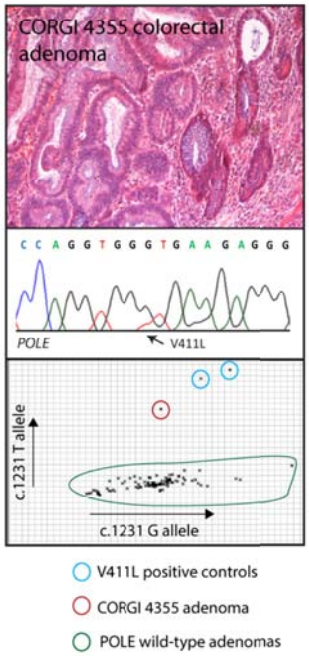


Figure 2. Whole genome sequencing of cancers with *POLE* exonuclease domain mutations

(A) Mutation burden and single nucleotide variant (SNV) type determined by whole genome sequencing (WGS) of five endometrial cancers (EC– Oxf001, POLE_040, POLE_049, POLE 072, POLE_147) and one colorectal cancer (CRC – Bir001) with somatic *POLE*^{P286R} exonuclease domain mutations. (B) Relative proportion of SNV mutations according to trinucleotide context averaged across the six *POLE*-mutant cases. The upper panel shows the unscaled proportions across the whole genome, while the lower panel shows the inferred mutational signature in a hypothetical genome where all trinucleotide frequencies are represented in equal proportions. High resolution versions are provided in supplementary material, Figure S1 (C) Frequency histograms and kernel density plots showing variant allele fraction (VAF) of all SNV mutations, and SNVs likely due to *POLE* exonuclease domain mutation (*POLE*). *POLE* mutations and other driver gene mutations are highlighted by arrows (details provided in supplementary material, Table S8). Only mutations in diploid regions of autosomes, and with coverage >20x are shown. The relatively low proportion of SNVs categorised as being due to *POLE* mutation reflects the stringency of the classification used (see Materials and methods, Mutational signatures). Vertical red line indicates clonal peak used to calculate cellularity.

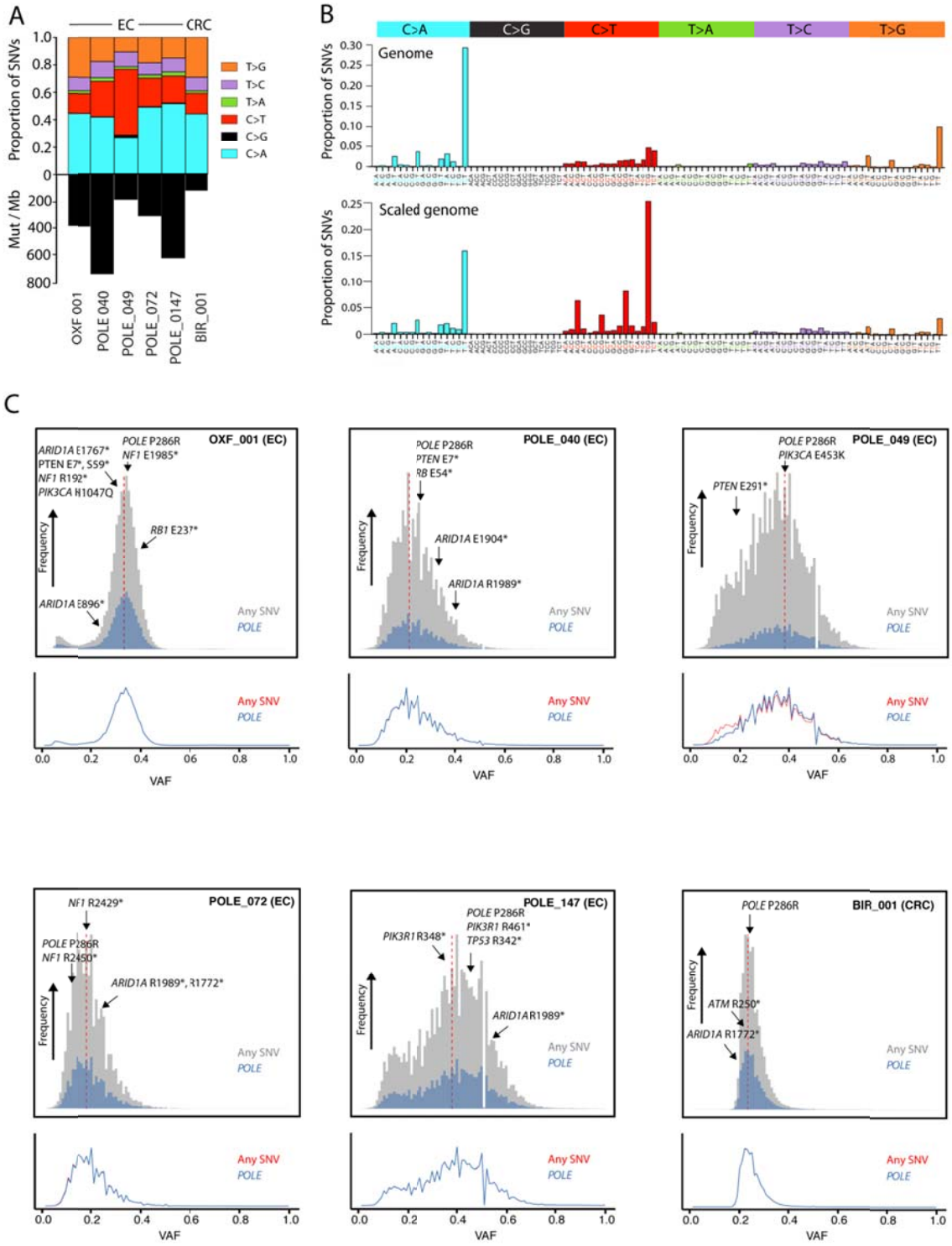


Figure 3. *POLE* signature mutations in endometrial cancer driver genes

Heatmap showing modelled probability that mutations in endometrial cancer driver genes (defined based on IntOGen – see Materials and methods, Definition of driver genes; supplementary material, Table S5) were due to a prior *POLE* exonuclease domain mutation. Results are shown for samples with a pathogenic *POLE* mutation and MMR-D and MMR-P comparators. Each non-synonymous mutation in a driver gene was assigned a probability that it was caused by the mutational process that generates the distinct *POLE* mutational signature, rather than by the mutational processes responsible for the consensus mutational signatures of *POLE*-wild-type DNA mismatch repair proficient (MMR-P) and mismatch repair deficient (MMR-D) tumours (see Materials and methods, *POLE* consensus mutational signature scores in driver genes, for details. For each gene/sample combination, a '*POLE*-score' was then calculated as the minimum value of these ratios, and plotted as a heatmap. Scores are shown for both individual *POLE*-mutant tumours and the combined *POLE*-mutant subgroup; results for tumours within the *POLE*-wild-type, mismatch repair proficient (MMR-P) and *POLE*-wild-type, mismatch repair deficient (MMR-D) subgroups are combined for clarity. Scores for *POLE* itself are shown for reference. Details of mutations are provided in supplementary material, Tables S8, S9. A high resolution version of this figure is provided as supplementary material, Figure S4.

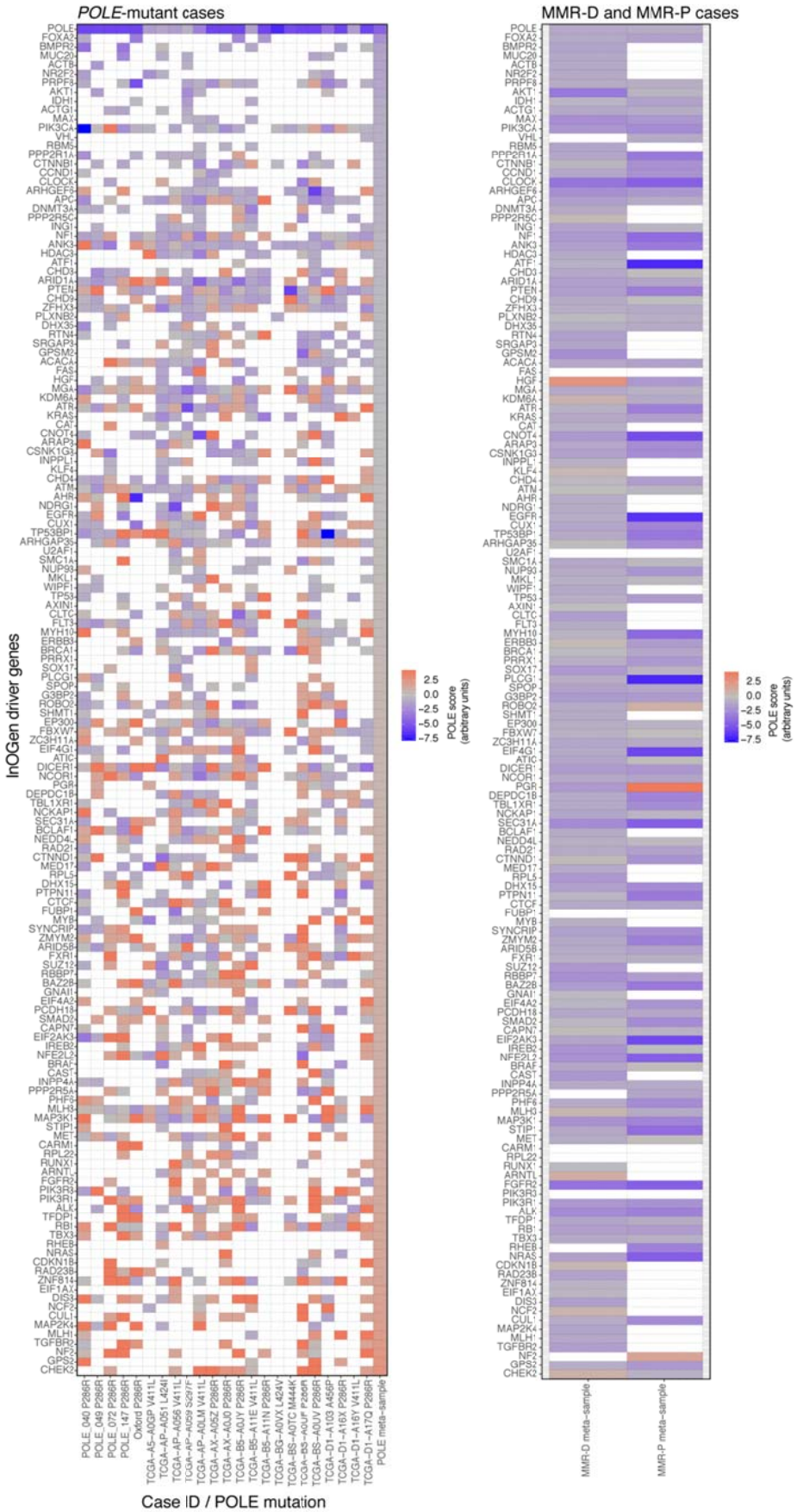


Figure 4. *POLE* signature mutations in colorectal cancer driver genes

Corresponding heatmap to Figure 3 showing results for known colorectal cancer driver genes, (defined base on IntOGen – see Materials and methods, Definition of driver genes; supplementary material, Table S4). Details of mutations are provided in supplementary material, Tables S8, S10. A high resolution version of this figure is provided as supplementary material, Figure S5.

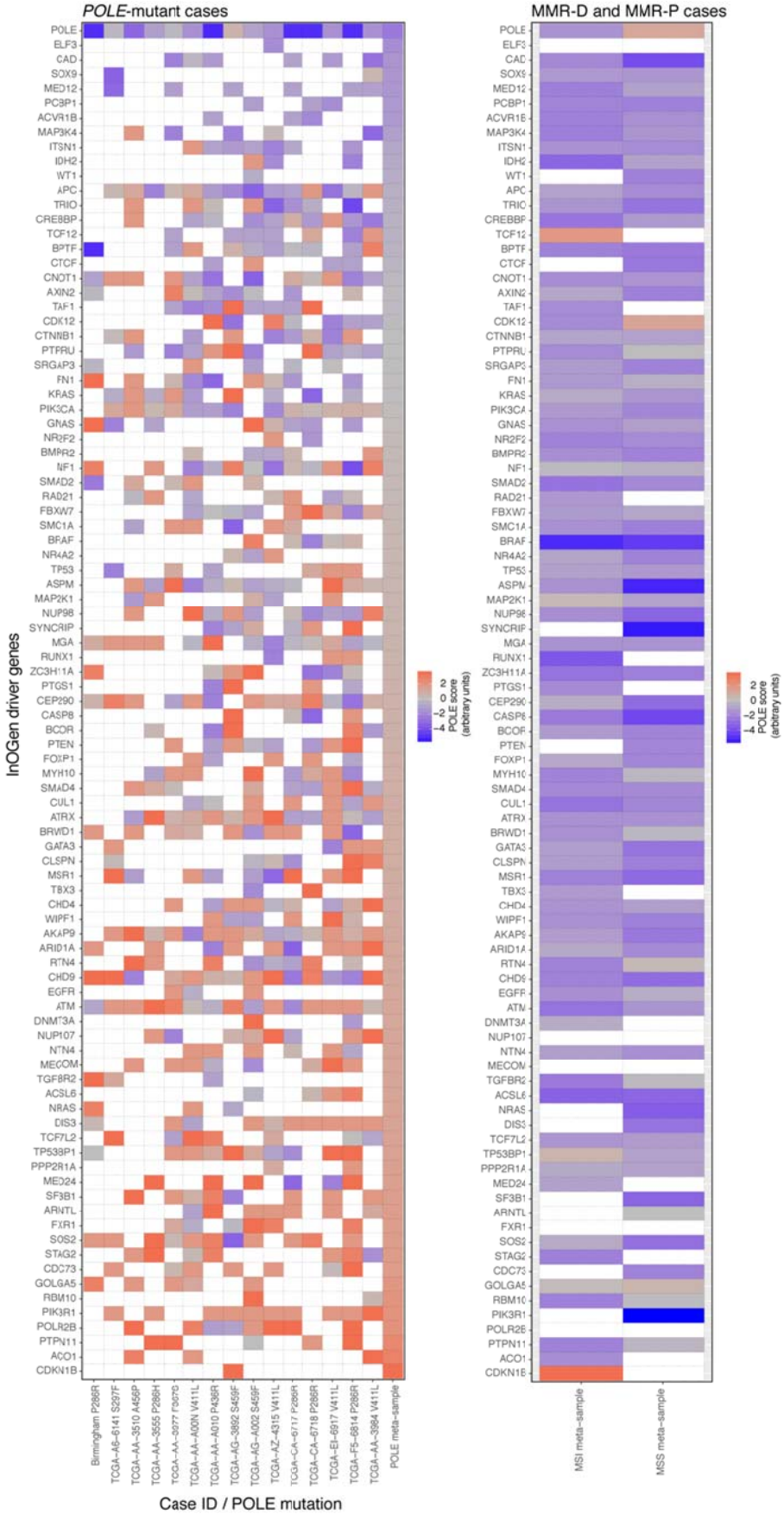


Figure 5. T cell infiltrate in *POLE*-mutant precursor lesions

(A) Representative images of immunohistochemistry (IHC) for the cytotoxic T cell marker CD8 in endometrial intraepithelial neoplasias (EIN) and paired concomitant endometrioid adenocarcinomas and in colorectal adenomas according to *POLE* mutation status. (B) Quantification of $CD8^+$ infiltrate density (number of $CD8^+$ cells per high power field [HPF] calculated as the mean of 10 high power fields) in *POLE*-wild-type and *POLE*-mutant paired endometrial intraepithelial neoplasia (EIN) and endometrial carcinoma (EC) (n=4 EIN–carcinoma pairs for each genotype) and in *POLE*-wild-type and *POLE*-mutant colorectal adenomas (Ad) (n=5 *POLE*-wild-type lesions, and the single *POLE*-mutant adenoma informative for analysis). Symbols (square, circle, triangle and diamond) correspond to paired EIN and endometrial carcinomas for *POLE*-wild-type (open symbols) and *POLE*-mutant (closed symbols) cases. For colorectal adenomas open and closed triangles correspond to unpaired *POLE*-wild-type and *POLE*-mutant adenomas respectively. Statistical comparisons in (B) were performed by unadjusted Mann-Whitney U-test. HPF – high power field.

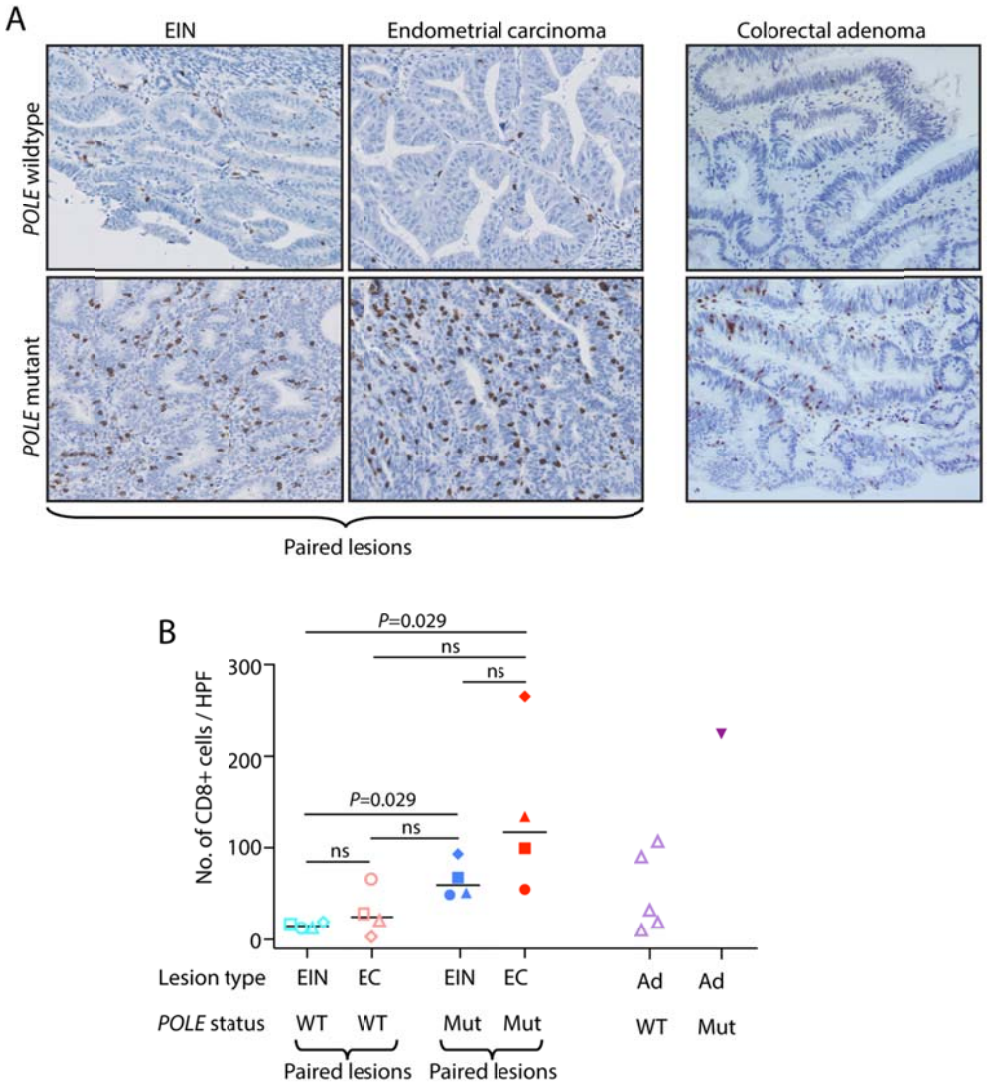
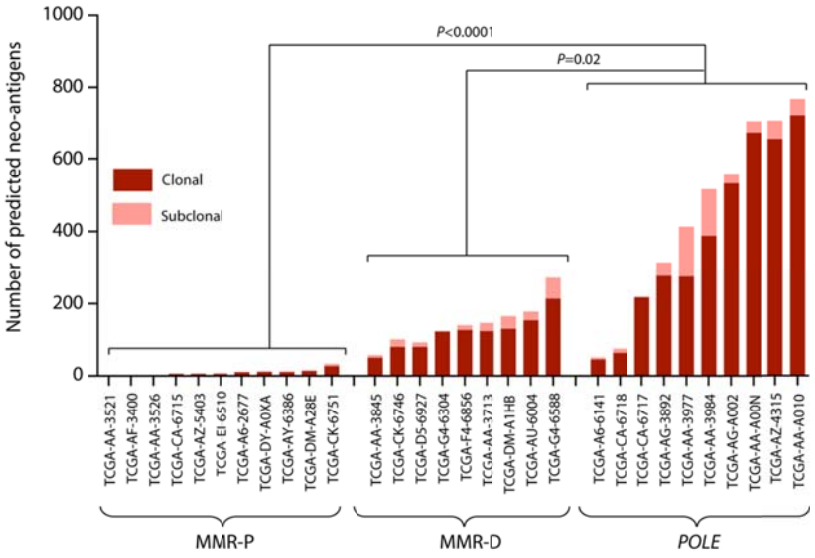


Figure 6. Clonality of predicted neoantigens in *POLE*-mutant colorectal cancers

Neoantigens were predicted based on the binding affinity of mutant peptides for patient class I HLA molecules, and assigned clonal or subclonal status (see Materials and methods, Clonality of *POLE* mutations). The number of clonal and subclonal neoantigens for *POLE*-wild-type, mismatch repair proficient (MMR-P), *POLE*-wild-type, mismatch repair deficient (MMR-D) and *POLE* mutant colorectal cancers from the TCGA series is shown. Cases in each molecular subgroup were selected to provide broadly similar proportions of disease stages and patient ages: molecular subgroups did not differ significantly in either parameter. Comparison of the number of clonal neoantigen burden between groups was performed by unadjusted Mann-Whitney U-test.



SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods YES

Supplementary figure legends YES

Table S6. List of IntOGen colorectal cancer driver genes used in this study

Table S7. Driver mutations detected in paired endometrial intraepithelial neoplasias (EIN) and endometrial carcinomas

Table S8. Driver mutations in *POLE*-mutant cancers analysed by whole genome sequencing

Table S9. Driver mutations in TCGA endometrial cancers by tumour molecular subgroup

Table S10. Driver mutations in TCGA colorectal cancers by tumour molecular subgroup

Table S11. Driver mutations in endometrial cancers analysed by Ion Ampliseq Comprehensive Cancer Panel